# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| 05/08/2012 | Final Technical Report | 04-01-2008 - 03-31-2012 |

**4. TITLE AND SUBTITLE**
Algorithms for Collective Knowledge Acquisition

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
N00014-08-1-0670

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Pedro Domingos

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
University of Washington
Office of Sponsored Programs
4333 Brooklyn Ave NE
Seattle, WA 98195-9472

**8. PERFORMING ORGANIZATION REPORT NUMBER**
UW budget 61-6756
UW eGC1 A38619

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Office of Naval Research
875 N Randolph Street
Arlington, VA 22203-1995

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
Public distribution/availability

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Collective knowledge bases (CKBs) allow knowledge from a multitude of sources to be efficiently gathered, integrated and deployed. Markov logic provides a representation language and learning and inference algorithms for CKBs. The goal of this project was to develop solutions to several problems that need to be addressed before CKBs can be widely deployed, including: tightly coupling learning and inference, learning many levels of structure, finding complex mappings across representations, optimizing joint inference, explaining the results of inference, and accepting natural language input.

**15. SUBJECT TERMS**
Knowledge Base, Markov Logic, Probability, Algorithms, Artificial Intelligence, Machine Learning, Computer Science.

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Pedro Domingos |
| | | | | | 19b. TELEPHONE NUMBER *(Include area code)* |
| UU | UU | UU | UU | 9 | 206-543-4229 |

# Final Report: N00014-08-1-0670
# Report Period: April 1, 2008 – March 31, 2012
# Algorithms for Collective Knowledge Acquisition


| | |
|---|---|
| Program Officer: | Behzad Kamgar-Parsi |
| Sponsoring Organization: | Office of Naval Research |
| Address: | 875 N Randolph Street, Suite 1425 |
| City, State, Zip: | Arlington, VA 22203-1995 |

| | |
|---|---|
| Principal Investigator: | Pedro Domingos |
| Performing Organization: | University of Washington |
| Address: | Computer Science & Engineering, Box 352350 |
| City, State, Zip: | Seattle, WA 98195-2350 |
| Phone Number: | 206-543-4229 |
| Fax Number: | 206-543-2969 |
| Email: | pedrod@cs.washington.edu |
| Website: | http://www.cs.washington.edu/homes/pedrod/ |


## Technical Section

Collective knowledge bases (CKBs) allow knowledge from a multitude of sources to be efficiently gathered, integrated and deployed. Markov logic provides a representation language and learning and inference algorithms for CKBs. The goal of this project was to develop solutions to several problems that need to be addressed before CKBs can be widely deployed, including: tightly coupling learning and inference, learning many levels of structure, finding complex mappings across representations, optimizing joint inference, explaining the results of inference, and accepting natural language input.

- We developed Tractable Markov Logic (TML), an expressive subset of Markov logic where inference is always tractable, and which therefore potentially makes reliable large-scale probabilistic knowledge bases feasible for the first time.

- We developed sum-product networks, a representation for propositional probabilistic models where inference is always tractable, and learning algorithms for it. (This work won the Best Paper Award at UAI-11.)

- We developed the first fully general probabilistic theorem proving techniques, building on our work on formula-based inference.

- We developed approximation by quantization, a new approach to probabilistic inference based on quantizing the levels of a potential when it becomes too large.

- We developed a new algorithm with theoretical guarantees for learning tractable Markov networks.

- We developed a new class of methods for approximate probabilistic inference, based on compilation of models to arithmetic circuits.

- We extended USP (our system for extracting knowledge from text without supervision) to build and populate ontologies, by clustering extracted objects and relations into multiple levels instead of just one.

- We developed new formula-based approach to MLN inference. This greatly scales up inference by eliminating and conditioning on formulas, instead of variables. Our system was a co-winner of the 2010 UAI Approximate Inference Challenge.

- We scaled up MLN inference and learning by using "coarse to fine" techniques, which take advantage of an ontology/type hierarchy to successively refine the inference while minimizing computational cost and maximizing approximation accuracy.

- We developed online inference techniques for MLNs.

- We developed approximate lifting techniques for MLN inference.

- We developed a new approach to learning MLN structure, using random walks to greatly improve efficiency.

- We developed the first bottom-up approach to learning Markov network structure.

- We published a book on Markov logic, synthesizing the research of the last few years.

- We developed an MLN and learning and inference algorithms for unsupervised semantic parsing. Our USP system extracts a knowledge base of logical formulas from raw text without any annotation or supervision, and uses it to answer questions. USP generalizes our earlier approach to unsupervised coreference resolution by resolving arbitrary constituents (not just noun phrases), and doing so compositionally. (This work won the Best Paper Award at EMNLP-09.)

- We generalized our lifted belief propagation approach to a wide variety of inference algorithms (e.g., variable elimination, junction trees, satisfiability, etc.), and developed an approximate version of it that further improves efficiency.

- We developed LHL, an algorithm for learning MLNs over entities and relations identified using our earlier SNE algorithm.

- We extended Markov logic to incorporate decision theory, and developed an algorithm for choosing the actions that maximize expected utility.

- We developed a state-of-the-art approach for unsupervised coreference resolution in text using Markov logic. This is one of the first unsupervised approaches to coreference, and the first one to match the accuracy of supervised approaches on standard corpora.

- We developed SNE, a system for extracting semantic networks from the Web based on Markov logic. SNE takes candidate tuples extracted by TextRunner, jointly clusters the objects and relations in them, and outputs the resulting semantic network (concepts and relations among them). This is the first system of its kind.

- We developed a series of efficient algorithms for inference in Markov logic.

## Awards and Honors

- Hoifung Poon received a Ph.D. degree in summer 2011 for his dissertation “Markov Logic in Machine Reading”.

- Best Paper Award at UAI 2011 for “Sum-Product Networks: A New Deep Architecture,” Hoifung Poon and Pedro Domingos, authors.

- Pedro Domingos was elected a Fellow of the Association for the Advancement of Artificial Intelligence in 2010 for significant contributions to the field of machine learning and to the unification of first-order logic and probability.

- Vibhav Gogate and Pedro Domingos won the PR track (estimating the partition function and probability of evidence) and placed second in the MAR track (marginal estimation) at the 2010 Uncertainty in Artificial Intelligence Approximate Inference Challenge.

- Stanley Kok received a Ph.D. degree in spring 2010 for his dissertation “Structure Learning in Markov Logic Networks”.

- Daniel Lowd received a Ph.D. degree in winter 2010 for his dissertation “Efficient Learning and Inference in Rich Statistical Representations”.

- Best Paper Award at EMNLP 2009 for “Unsupervised Semantic Parsing,” Hoifung Poon and Pedro Domingos, authors.

- Invited column, “Solving AI: We Need a New Language for Artificial Intelligence,” by Pedro Domingos in *Technology Review*, March/April 2009 Vol.112/No.2.

- Parag Singla received a Ph.D. degree in winter 2009 for his dissertation “Markov Logic: Inference, Learning, Applications, and Extensions”.

## Patent Filings or Patent Awards

There were no patent filings or patent awards resulting from this grant.

## Publications


1. William Austin Webb and Pedro Domingos, “A Tractable First-Order Probabilistic Logic,” Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012. Toronto, Canada: AAAI Press. To appear.

2. Hoifung Poon and Pedro Domingos, “Sum-Product Networks: A New Deep Architecture,” Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (pp. 337-346), 2011. Barcelona, Spain: AUAI Press. Winner of the Best Paper Award.

3. Vibhav Gogate and Pedro Domingos, “Probabilistic Theorem Proving,” Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (pp. 256-265), 2011. Barcelona, Spain: AUAI Press.

4. Vibhav Gogate and Pedro Domingos, “Approximation by Quantization,” Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (pp. 247-255), 2011. Barcelona, Spain: AUAI Press.

5. Chloe Kiddon and Pedro Domingos, “Coarse-to-Fine Inference and Learning for First-Order Probabilistic Models,” Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (pp. 1049-1056), 2011. San Francisco, CA: AAAI Press.

6. James Blythe, Jerry Hobbs, Pedro Domingos, Rohit Kate and Raymond Mooney, “Implementing Weighted Abduction in Markov Logic,” Proceedings of the Ninth International Conference on Computational Semantics (pp. 55-64), 2011. Oxford, UK, 2011: ACL SIGSEM.

7. Vibhav Gogate, William Austin Webb and Pedro Domingos, “Learning Efficient Markov Networks,” Advances in Neural Information Processing Systems 23 (pp. 748-756), 2010. Red Hook, NY: Curran Associates.

8. Daniel Lowd and Pedro Domingos, “Approximate Inference by Compilation to Arithmetic Circuits,” Advances in Neural Information Processing Systems 23 (pp. 1477-1485), 2010. Red Hook, NY: Curran Associates.

9. Hoifung Poon and Pedro Domingos, “Unsupervised Ontology Induction from Text,” Proceedings of the Forty-Eighth Annual Meeting of the Association for Computational Linguistics (pp. 296-305), 2010. Uppsala, Sweden: ACL.

10. Vibhav Gogate and Pedro Domingos, “Formula-Based Probabilistic Inference,” Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (pp. 210-219), 2010. Catalina Island, CA: AUAI Press.

11. Aniruddh Nath and Pedro Domingos, “Efficient Belief Propagation for Utility Maximization and Repeated Inference,” Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (pp. 1187-1192), 2010. Atlanta, GA: AAAI Press.

12. Aniruddh Nath and Pedro Domingos, “Efficient Lifting for Online Probabilistic Inference,” Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (pp. 1193-1198), 2010. Atlanta, GA: AAAI Press.

13. Vibhav Gogate and Pedro Domingos, “Exploiting Logical Structure in Lifted Probabilistic Inference,” AAAI 2010 Workshop on Statistical Relational AI, 2010.

14. Chloe Kiddon and Pedro Domingos, “Leveraging Ontologies for Lifted Probabilistic Inference and Learning,” AAAI 2010 Workshop on Statistical Relational AI, 2010.

15. Hoifung Poon and Pedro Domingos, “Machine Reading: A 'Killer App' for Statistical Relational AI,” AAAI 2010 Workshop on Statistical Relational AI, 2010.

16. Parag Singla, Aniruddh Nath and Pedro Domingos , “Approximate Lifted Belief Propagation,” AAAI 2010 Workshop on Statistical Relational AI, 2010.

17. Jesse Davis and Pedro Domingos, “Bottom-Up Learning of Markov Network Structure,” In Proceedings of the Twenty-Seventh International Conference on Machine Learning (pp. 271-278), 2010. Haifa, Israel: Omnipress.

18. Stanley Kok and Pedro Domingos, “Learning Markov Logic Networks Using Structural Motifs,” Proceedings of the Twenty-Seventh International Conference on Machine Learning (pp. 551-448), 2010. Haifa, Israel: Omnipress.

19. Hoifung Poon and Pedro Domingos, “Unsupervised Semantic Parsing,” Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (pp. 1-10), 2009. Singapore: ACL. Winner of the Best Paper Award.

20. Jesse Davis and Pedro Domingos, “Deep Transfer via Second-Order Markov Logic,” Proceedings of the Twenty-Sixth International Conference on Machine Learning (pp. 217-224), 2009. Montreal, Canada: Omnipress.

21. Stanley Kok and Pedro Domingos, “Learning Markov Logic Network Structure via Hypergraph Lifting,” Proceedings of the Twenty-Sixth International Conference on Machine Learning (pp. 505-512), 2009. Montreal, Canada: Omnipress.

22. Aniruddh Nath and Pedro Domingos, “A Language for Relational Decision Theory,” Proceedings of the Sixth International Workshop on Statistical Relational Learning, 2009. Leuven, Belgium.

23. Hoifung Poon and Pedro Domingos, “Joint Unsupervised Coreference Resolution with Markov Logic,” Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (pp. 649-658), 2008. Honolulu, HI: ACL.

24. Pedro Domingos and Daniel Lowd, “Markov Logic: An Interface Layer for Artificial Intelligence,” San Rafael, CA: Morgan & Claypool, 2009.

25. Hoifung Poon and Pedro Domingos, “Joint Unsupervised Coreference Resolution with Markov Logic,” in Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (pp. 650-659), 2008. Waikiki, Hawaii: ACL/Omnipress.

26. Stanley Kok and Pedro Domingos, “Extracting Semantic Networks from Text via Relational Clustering”, in Proceedings of the Nineteenth European Conference on Machine Learning (pp. 624-639), 2008. Antwerp, Belgium: Springer.

27. Daniel Lowd and Pedro Domingos, “Learning Arithmetic Circuits,” in Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (pp. 383-392), 2008. Helsinki, Finland: AUAI Press.

28. Parag Singla and Pedro Domingos, “Lifted First-Order Belief Propagation,” in Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (pp. 1094-1099), 2008. Chicago, IL: AAAI Press.

29. Jue Wang and Pedro Domingos, “Hybrid Markov Logic Networks,” in Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (pp. 1106-1111), 2008. Chicago, IL: AAAI Press.

30. Hoifung Poon, Marc Sumner and Pedro Domingos, “A General Method for Reducing the Complexity of Relational Inference and its Application to MCMC,” in Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (pp. 1075-1080), 2008. Chicago, IL: AAAI Press.

31. Jesse Davis and Pedro Domingos, “Deep Transfer via Second-Order Markov Logic,” in Proceedings of the AAAI-2008 Workshop on Transfer Learning for Complex Tasks, 2008. Chicago, IL: AAAI Press.

## Presentations

Invited talks by Pedro Domingos:

1. Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (Barcelona, Spain; invited tutorial, with Kristian Kersting), 2011.

2. Tenth International Conference on Machine Learning and Applications (Honolulu, HI; invited tutorial), 2011.
3. ICML-2011 Workshop on Learning Architectures, Representations, and Optimization for Speech and Visual Information Processing (Bellevue, WA), 2011.

4. Summer School on Probabilistic Models of Cognition (Los Angeles, CA), 2011.

5. University of Texas, Austin, 2011.

6. University of Memphis (Memphis, TN), 2011.

7. Swiss Federal Institute of Technology (ETH; Zurich, Switzerland), 2011.

8. Google Faculty Summit (Zurich, Switzerland), 2011.

9. Los Alamos National Laboratory (Los Alamos, NM), 2010.

10. Georgia Institute of Technology (Atlanta, GA), 2010.

11. IBM Thomas J. Watson Research Center (Yorktown Heights, NY), 2010.

12. Twenty-Fifth Snowbird Learning Workshop (Snowbird, UT), 2010.

13. Sixth International Workshop on Neural-Symbolic Learning and Reasoning (Atlanta, GA), 2010.

14. Seattle Robotics Society (Renton, WA), 2010.

15. University of Edinburgh (Scotland), 2009.

16. Johns Hopkins University (Baltimore, MD), 2009.

17. MITACS 2009 Annual Conference (Fredericton, Canada).

18. Re: Learning Conference (Washington, DC), 2009.

19. Workshop on Information in Networks (New York University), 2009.

20. CALO 2009 Annual Meeting (SRI International, Menlo Park, CA).

21. Carnegie Mellon University (Pittsburgh, PA), 2008.

22. Cornell University (Ithaca, NY), 2008.

23. University of Granada (Spain), 2008.

24. Defense Advanced Research Projects Agency (Arlington, VA), 2008.

25. Twenty-Fourth International Conference on Logic Programming (Udine, Italy), 2008.

26. Seventeenth ACM Conf. on Information and Knowledge Management (Napa Valley, CA), 2008.

27. Twelfth International Workshop on Structural and Syntactic Pattern Recognition and Seventh International Workshop on Statistical Techniques in Pattern Recognition (joint invited talk; Orlando, FL), 2008.

28. ICML-2008 Workshop on Prior Knowledge for Text and Language (Helsinki, Finland).

29. NIPS-2008 Workshop on Speech and Language: Learning-based Methods and Systems (Whistler, Canada).

30. NIPS-2008 Workshop on Beyond Search: Computational Intelligence for the Web (Whistler, Canada).

31. ONR Workshop on Research Directions in Information Integration (Monterey, CA), 2008.

## People Supported

1. Pedro Domingos, Principal Investigator

2. Robert Gens, Graduate Student

3. Aniruddh Nath, Graduate Student

4. Hoifung Poon, Graduate Student

## Project and Related Websites

http://alchemy.cs.washington.edu/
http://ai.cs.washington.edu/
http://www.cs.washington.edu/homes/pedrod/